DOI:
10.1007/s10519-012-9561-y

Citation for published version (APA):
Pinto, R., Rijsdijk, F., Frazier-Wood, A. C., Asherson, P., & Kuntsi, J. (2012). Bigger families fare better: a novel method to estimate rater contrast effects in parental ratings on ADHD symptoms. Behavior Genetics, 42(6), 875-885. DOI: 10.1007/s10519-012-9561-y

Published in final edited form as:
Behav Genet. 2012 November ; 42(6): 875–885. doi:10.1007/s10519-012-9561-y.

Bigger families fare better: a novel method to estimate rater contrast effects in parental ratings on ADHD symptoms

R. Pinto[1], **F. Rijsdijk**[1], **A. C. Wood**[1,2], **P. Asherson**[1], and **J. Kuntsi**[1]

[1]King's College London, MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, London, UK

[2]Department of Epidemiology and Section on Statistical Genetics, University of Alabama at Birmingham, Birmingham, Alabama, USA

Abstract

Many twin studies on parental ratings of attention deficit hyperactivity disorder (ADHD) symptoms report low or negative DZ correlations. The observed differences in MZ and DZ variances indicate sibling contrast effects, which appear to reflect a bias in parent ratings. Knowledge of the factors that contribute to this rater contrast effect is, however, limited. Using parent-rated ADHD symptoms from the Twins' Early Development Study and a novel application of a twin model, we explored a range of socio-demographic variables (ethnicity, socio-economic status, and family size), as potential contributors to contrast effects and their interactive effect with gender composition of twin pairs. Gender did moderate contrast effects but only in DZ opposite-sex twin pairs. Family size also showed a moderating effect on rater contrast effects, which was further modified by gender. We further observed an effect of rating scale, with the DSM-IV ADHD subscale of the Revised Conners' Parent Rating Scale more resistant to contrast effects than shorter scales of ADHD symptoms. The improved identification of situations where the accuracy of the most common informant of childhood ADHD symptoms – parents – is compromised as a result of rater bias, might have implications for future research on ADHD.

Keywords

ADHD; twin study; contrast effects; rater bias

INTRODUCTION

Attention deficit hyperactivity disorder (ADHD) is characterised by developmentally inappropriate levels of hyperactivity, impulsivity and inattention, and is one of the most frequently diagnosed childhood-onset disorders. Pooled quantitative genetic studies yield mean heritability estimates of 0.70 to 0.76 (Burt 2009; Faraone et al. 2005), making ADHD one of the most heritable psychiatric disorders (Plomin et al. 2008). However, rater contrasts in parental ratings artificially amplify differences between MZ and DZ correlations, yielding potentially inflated heritability estimates (Freitag et al. 2010; Wood et al. 2010). In line with

Correspondence to Rebecca Pinto, King's College London, MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, London, UK. rebecca.pinto@kcl.ac.uk, Telephone: +44 (0)207 848 5415, Fax: +44 (0)207 848 0866.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

this, a review reported lower heritability estimates for teacher-rated ADHD symptoms (around 50%), and objective (actigraph-measured) measures of activity level (30–52%) (Freitag et al. 2010).

The twin design can disentangle genetic and environmental influences underlying variation of familial disorders, based on differences in genetic relatedness between monozygotic (MZ; identical) twins, who share all their genetic variation, and dizygotic (DZ; non-identical) twins, who share on average 50% of their genetic variation. Several twin studies on parental ADHD ratings have reported low DZ correlations that are less than half the MZ twin correlations, indicating potential genetic dominance (Kuntsi et al. 2004; Price et al. 2005; Rietveld et al. 2003; Saudino et al. 2005). Low DZ correlations in tandem with significantly larger DZ variances (relative to MZ), indicate contrast effects, which may be attributed to either competitive sibling interaction, whereby the behavior of one twin influences the behavior of the co-twin (reflecting true phenotypic differences), or a form of rater bias where parents emphasize behavioral differences. The evidence suggests parental rater bias (Simonoff et al. 1998), supported by the lack of very low DZ correlations for objective measures of ADHD-related behaviors (Saudino et al. 2000; Saudino et al. 2004) and teacher ADHD ratings (Saudino et al. 2005; Martin et al. 2002). The absence of rater contrast effects in teacher ratings are hypothesized to stem from greater exposure to children and childhood behavioral norms (Hartman et al. 2007).

ADHD ratings seem particularly susceptible to parental rater contrast effects (Saudino et al. 2005), perhaps because they are based on more subjective criteria. Parental ratings of other behavioral traits, such as conduct problems, show no contrast effects (Saudino et al. 2005; Hudziak et al. 2005), potentially attributable to increased awareness of clearly defined societal norms for more socially disruptive behaviors (Simonoff et al. 1998).

However, contrast effects are not universally found in parental ratings of ADHD, but vary according to a number of (potentially interacting) factors. A first factor to consider is the psychometric properties of the rating scale (Thapar et al. 2000). Overall, contrast effects seem less likely to arise when rating scales contain specific descriptions of behavior (Saudino et al. 2004), are longer and more detailed (Kuntsi et al. 2005) such as DSM-IV symptom checklists, and in scales with a broader scoring range (Hay et al. 2007; Polderman et al. 2007).

Another factor that might influence rater contrast effects of parental ADHD ratings is developmental stage of the behavior. A previous study based on the present sample (Twins' Early Development Study, TEDS) reported rater bias in parental ADHD ratings in children aged 2, 3 and 4, and concluded that contrast effects contributed to the stability of ADHD symptom ratings in pre-school children (Price et al. 2005). In line with this, a study based on a sub-sample of TEDS (Kuntsi et al. 2004) found contrast effects in DSM-IV ADHD ratings at age 5. Follow-up studies extending to middle-childhood and early adolescence found no contrast effects in ratings obtained from the Revised Conners' Parent Rating Scale at age 8 (Kuntsi et al. 2005; McLoughlin et al. 2007) or 12 (Greven et al. 2011), which may, of course, also be attributable to changes in assessment instrument. Within the TEDS sample, comparisons of parental ratings from the Strengths and Difficulties Questionnaire for twins aged 12 (Merwood et al. Submitted), and pre-school children (Price et al. 2005; Price et al. 2001), suggest a developmental decline in the magnitude of parental contrast effects. This pattern of results has also been found in an independent twin sample for maternal Child Behavior Checklist (CBCL) ADHD-related ratings, which the authors hypothesized could be attributable to parents increased exposure to children as they get older (Rietveld et al. 2003).

Third, gender may influence rater contrast effects of parental ADHD ratings. One study found evidence of contrast effects in parental hyperactivity-impulsivity ratings, but this was limited to females (Vierikko et al. 2004). In contrast, another two studies on independent twin samples found that contrast effect parameters could be equated within same-sex pairs, suggesting that – when parents rate one twins behavior in relation to their co-twin – this was not moderated by gender (Rietveld et al. 2003; Simonoff et al. 1998). Moreover, these two studies showed that contrast effects could be further constrained between same-sex and opposite-sex pairs, suggesting that even in cases when twin members are not of the same gender, they are compared to a similar extent as same-sex pairs (Rietveld et al. 2003; Simonoff et al. 1998). Yet another study reported significantly larger contrast effects for same-sex pairs, suggesting that when twin members of a pair differ by gender, ratings are more independent and less influenced by rater bias (van Beijsterveldt et al. 2004). Only one of these studies has further decomposed contrast effect parameters within DZ opposite-sex pairs, estimating a contrast effect parameter from males-to-females (the female member is evaluated in relation to their male co-twin) and from females-to-males (the male member is evaluated in relation to their female co-twin) (Rietveld et al. 2003). Within opposite-sex pairs, rater contrast parameters did differ by gender, with a larger effect observed from males-to-females, compared to from females-to-males, suggesting that when opposite-sex pairs are being rated for ADHD-related behaviours, the male twin is considered the standard and the female twin evaluated accordingly (Rietveld et al. 2003). The authors hypothesized that the tendency to use males as a comparative benchmark may stem from ADHD-related behaviors being more commonly associated as male traits and more frequently observed in males.

The aim of this paper is to explore potential explanatory factors for contrast effects, employing a novel non-genetic model; therefore we do not investigate aetiological components of ADHD. Few studies have investigated additional factors that may influence contrast effects. One study reported differences in parental extraversion, but not socio-economic status (SES) or education, as contributing to contrasting non-twin siblings (Saudino et al. 2004). Although sibling-pair constellation variables such as number of children, sex composition, age distribution have been hypothesized as contributing to the tendency to contrast children (Carey 1986), differences in gender composition and age were not correlated with behavioural difference scores (indexing contrast effects) in non-twin siblings (Saudino et al. 2004). Family size has not been formally examined as moderating the process of contrasting twins. In this study we examine several parent- and child-related socio-demographic characteristics as potential factors that may contribute to contrast effects. Specifically, we present a model which will enable us to: (1) explore if contrast effects in parental ADHD ratings differ according to gender composition of rated twins, and (2) estimate the interactive role of gender with other demographic factors (ethnicity, socio-economic status and family size) that may contribute to parental rater contrast effects, using

Europe PMC Funders Author Manuscripts

repeated measures of the Revised Parent Rutter Scale for Pre-School Children (RRPSPS) (Hogg et al. 1997), Strengths and Difficulties Questionnaire (SDQ) (Goodman 1997), and the Revised Conners' Parent Rating Scale (CPRS-R) (Conners et al. 1998), at a number of time points from early childhood to pre-adolescence. The improved identification of situations where the accuracy of the most common informant of childhood ADHD symptoms – parents – is compromised, might contribute to our theoretical understanding of this puzzling effect.

METHODOLOGY

Sample

Participants are members of the Twins' Early Development Study (TEDS) (Trouton et al. 2002), a population-based birth cohort of twins born in 1994-1996. All families in England and Wales identified by the Office for National Statistics as having twins born in these years were invited to enrol when the twins were aged 18 months old. Parents of all participants have provided informed consent and the study has been approved by the Institute of Psychiatry Ethical Committee (approval number 183/94). The 18-month booklet contained questions relating to pregnancy, birth, and socio-demographic indicators. Zygosity status was initially assigned based on a standard parent-rated zygosity questionnaire that has been shown to have a greater than 95% accuracy rate, compared to zygosity determined by DNA testing (Price et al. 2000). Zygosity for the vast majority of the sample has been subsequently confirmed by the employment of DNA markers. Despite attrition and non-responses over time, TEDS families at each age remain reasonably representative of the UK population in terms of parental education, parental employment and ethnicity (Trouton et al. 2002).

Twin pairs were excluded from the current analysis if there were extreme pregnancy or perinatal difficulties, specific medical syndromes and chromosomal anomalies, or if first contact data or zygosity information was unavailable. After exclusion criteria, symptom scores using the Revised Parent Rutter Scale for Pre-School Children (RRPSPS) (Hogg et al. 1997) at ages 2, 3, and 4 were obtained for 9153, 9437, 12974 individual twins, respectively. At ages 4, 7 and 12 symptom scores were derived from the Strengths and Difficulties Questionnaire (SDQ) (Goodman 1997), for 12966, 14359, 11170 individual twins. At ages 8 and 12, ratings from the Revised Conners' Parent Rating Scale (CPRS-R) (Conners et al. 1998) were available for 12518 and 11181 individual twins.

Measures

ADHD symptoms—At the 2-, 3- and 4-year data collection points, parents were asked to rate the behavior of each twin using the Revised Parent Rutter Scale for Pre-School Children (RRPSPS) (Hogg et al. 1997). The rater reported on the frequency of behavioral attributes using a three-point scale: 0 indicates a response of "not true", 1 indicates "sometimes true" and 2 indicates "certainly true". The current analyses focused on four items that make up the hyperactivity-inattentive subscale ("restless; runs about or jumps up and down, doesn't keep still"; "squirmy, fidgety"; "has poor concentration, or short attention span"; and "inattentive"). The total ADHD symptom score was calculated by summing scores for each

Europe PMC Funders Author Manuscripts

rated item, requiring assessment of at least two items for a valid measurement, with a higher score inferring increased levels of ADHD-related behaviors. In the present sample the internal consistency reliability of the scale was .70 at age 2, .72 at age 3, and .73 at age 4.

Parental behavioral ratings from the Strengths and Difficulties Questionnaire (SDQ) (Goodman 1997) were obtained at the 4-, 7- and 12-year data collection points. The hyperactivity-inattention subscale of the SDQ is similar to the RRPSPC scale, in that it has a three-point scale, and contains three overlapping (but slightly differently worded) items. In total there are 5 items ("restless, overactive, cannot stay still for long"; "constantly fidgeting or squirming"; "easily distracted, concentration wanders"; "thinks things out before acting" (reversed); "see tasks through to the end, good attention span" (reversed)). The total ADHD symptom score was calculated from the total sum of scores for each rated item (requiring at least three of five items to be rated), with a higher score indicative of higher measures of ADHD symptoms. The internal consistency reliability of the scale was .76 at each of the 3 time-points (at ages 4, 7 and 12).

When twins were aged, on average, 8 and 12 years, parents were asked to rate the behavior of each twin using the Revised Conners' Parent Rating Scale (CPRS-R) (Conners et al. 1998). The CPRS-R has two DSM-IV symptom sub-scales (inattentiveness and hyperactivity-impulsivity), each consisting of nine items that map onto DSM-IV criteria. The scale uses a four-point scale and the total score on each sub-scale can be obtained by calculating the sum of all dimensional item scores (values between 0 and 27; requiring at least 5 of the 9 items). The sum of all 18 items (at least nine to be non-missing) calculates a total DSM-IV ADHD symptom score (values between 0 and 54), with a greater score indicating a greater rating of ADHD symptoms. The internal consistency reliability of the CPRS-R was .93 at both 8 and 12 years.

Socio-demographic factors

Child's gender: Gender was re-coded so that females, which constituted the largest proportion of the total sample, were assigned as the reference group (0) (see Table 1 for breakdown of socio-demographic variables by sex-specific zygosity groups).

Child's ethnicity: Parents were asked to nominate twins' ethnicity based on a choice of 5 broad ethnic categories that were used in the 1991 UK Census: White, Black, Asian, Other and Mixed, which have been shown to map onto the more detailed 16 ethnic categories used in the UK 2001 Census (Kumarapeli et al. 2006). The vast majority of the sample had parent-nominated ethnicity: only 22 twin pairs have no recorded ethnicity. Out of 8770 twin pairs with ethnicity data, 93% had their ethnicity assigned as White (n=8135 twin pairs); 3% assigned Mixed (n=262); 1.95% assigned Asian (n=171); 1.36% assigned Black (n=119); and 0.70% assigned Other (n=61). Small samples across minority ethnic groups led to them being collapsed to produce one category (n=613). Ethnicity categories were re-coded as 1 or 0, such that 0 indicated the group with the larger sample size. Accordingly the white group was coded as 0, and minority ethnic group coded as 1.

Socio-economic status (SES): SES was measured using demographic information collected at initial contact, and was missing for 11% of the entire sample (n=2076 individual twins).

An index of SES was created based on a factor analysis of maternal and paternal occupational status and highest educational attainment (Pike et al. 2006). Age of mother at the birth of her eldest child was also included as an indicator of low SES, as it loaded on the same factor. These five SES indicators were standardised, and then summed using unit weights in order to create a general single composite measure of SES, with a lower value representing a higher level of risk of low SES (Pike et al. 2006).

Family size: At the point of initial contact, when twins were aged 18 months and parents consented to participate in TEDS, parents were asked to provide information relating to the family. Parents were asked "how many other children live in the home with your twins?" and asked to break down the number of additional children by gender and by whether they were older or younger than twin pairs. Using this information, a variable relating to family size was created, to act as a proxy for exposure to children, based on the total number of additional children in the household.

Analyses

Model-Fitting procedure—In accordance with standard quantitative genetic procedures, transformations were applied to normalize the positively skewed distributions for ADHD symptom score ratings derived from the CPRS-R at ages 8 and 12, using the optimized minimal skew 'lnskew0' command in STATA (StataCorporation 1997). We then adopted a step-wise procedure to test for contrast effects.

Step 1: Testing variance differences according to zygosity to indicate possible contrast effects (twin correlation model without interaction parameters)

In the first series of models, before modeling contrast effects, MZ and DZ twin correlations and variance estimates for non-adjusted parental ratings were obtained. Significance of variance differences between MZ and DZ (including DZ opposite-sex) pairs was evaluated by likelihood ratio testing, comparing a sub-model where variances were constrained to be equal across zygosity (MZ versus DZ) to one in which they were freely estimated across zygosity. Variance inequality tests were run separately for males and females to determine if gender moderated rater contrast effects. If findings were not consistent with the presence of contrast effects, these variables were dropped from subsequent analyses.

Step 2: Testing if contrast effects significantly differ between same-sex and opposite-sex twin pairs according to gender (twin correlation model with overall interaction effects)

The rater contrast model—A non-genetic model was used to test rater contrast effects and the potential explanation of the socio-demographic factors on this effect. The model specified the variance-covariance structure of the MZ and DZ data as a $(I-B)^{inv} * (S*R*S') * ((I-B)^{inv})'$. The S*R*S' part is a Gaussian decomposition of the variance-covariance structure of the data, where S is a 2 × 2 diagonal matrix with the standard deviations of the twin 1 and twin 2 scores of the phenotype under study, regardless of twin order and zygosity but sex specific (S_M and S_F, Figure 1); and R is a 2 × 2 correlation matrix between twin 1 and twin 2 score, estimating just one for MZ pairs (constrained across males and females) and one for DZ pairs (constrained across same-sex and opposite-sex pairs). The model further allowed for sex specific means (μ_M, μ_F). These specifications are mainly based on

Europe PMC Funders Author Manuscripts

previous findings of the same sample indicating that in parental ADHD ratings there are no quantitative or qualitative sex differences, but consistent evidence for sex differences in variance (Price et al. 2005; Saudino et al. 2005; McLoughlin et al. 2007; Greven et al. 2011; Price et al. 2001).

The rater contrast part is modeled in the (I-B) structure and is a standard way of specifying reciprocal causation pathways between two internal variables in structural equation models to overcome infinite feedback loops, where I is a 2 × 2 identity matrix and B is a 2 × 2 matrix, with zeros on the diagonals and symmetric off-diagonal parameters representing the causal paths (the *c* paths in Figure 1).

In the second stage of analysis (disregarding the socio-demographic explanatory factors), four sex-by-zygosity rater contrast effect parameters were estimated to capture the possibility that different rater contrast effects may be present in pairs with varying composition of sex: male-male (c_m), female-female (c_f), male-to-female ($c_{M\text{-}F}$) and female-to-male ($c_{F\text{-}M}$). The power to detect c_m and c_f is due to the fact that in this model the predicted variances and covariances will differ across MZ and DZ same-sex pairs. The power to detect $c_{F\text{-}M}$ is based on observed differences in variance between same-sex males and DZ opposite-sex males. The power to detect $c_{M\text{-}F}$ is based on observed differences in variance between same-sex females and DZ opposite-sex females. Age and all other moderators (family size, SES, and ethnicity) were incorporated as covariates in the model of the means (effectively regressing out any confounding effects). A series of sub-models were run to test whether c_m and $c_{M\text{-}F}$ could be equated, and whether c_f and $c_{F\text{-}M}$ could be equated, for the final stage of analysis.

Step 3: Testing moderators of contrast effects (twin correlation model with independent and moderator-dependent interaction effects)

Using the full form of the rater contrast model (see figure 1), certain demographic variables of interest were incorporated to explore the extent to which they contributed to contrast effects. This involves splitting up the ‘total contrast effect’ (c_m, c_f, $c_{M\text{-}F}$, and $c_{F\text{-}M}$ in Figure 1) into a moderator-independent part (i_m, i_f, $i_{M\text{-}F}$, and $i_{F\text{-}M}$) and a moderator-dependent part (k_m, k_f, $k_{M\text{-}F}$, and $k_{F\text{-}M}$), which is estimated by means of moderators on the interaction paths, incorporated as definition variables in Mx. If findings from the second stage of analysis suggest that rater contrast parameters can be equated within gender groups (i.e. $c_m = c_{M\text{-}F}$; $c_f = c_{F\text{-}M}$), then only two sex-specific contrast effect parameters would be specified, and moderators included on these pathways. If not, four sex-by-zygosity contrast effect parameters would be specified (c_m, c_f, $c_{M\text{-}F}$, and $c_{F\text{-}M}$), and moderators modeled on each pathway. The power to detect c_m and c_f is due to the fact that in this model the predicted variances and covariances will differ across MZ and DZ same-sex pairs. The power to detect $c_{F\text{-}M}$ is based on observed differences in variance between same-sex males and DZ opposite-sex males. The power to detect $c_{M\text{-}F}$ is based on observed differences in variance between same-sex females and DZ opposite-sex females. Age was further incorporated as a covariate in the model of the means.

The structural equation-modeling program Mx (Neale 1997) was used to conduct the analyses. Participants with missing data were included in the analyses, as Mx provides a method for handling incomplete data by using raw maximum likelihood estimation, in which a likelihood statistic (−2LL) for each observation is calculated. This implies that there is no overall measure of fit. Instead, with raw data, there are relative measures of fit: by comparing the −2LL of nested models a chi-square goodness-of-fit index (χ^2) is obtained, relative to a change in degrees of freedom (df). We adopted a p value of 0.01, to control for multiple testing.

RESULTS

Step 1: Testing variance differences according to zygosity to indicate possible contrast effects (twin correlation model without interaction parameters)

Twin correlations and tests for variance differences (see Table 2) were examined to determine whether there was evidence of contrast effects in parental ratings. Twin correlations for all RRPSPS and SDQ ratings were consistent with the presence of rater contrast or non-additive genetic effects. Conversely, DZ correlations for CPRS-R ratings were greater than half of MZ correlations. Formal testing of zygosity differences in phenotypic variance by gender confirmed that contrast effects were not present for CPRS-R ratings. Consequently, CPRS-R ratings were not included in further analysis.

Evidence of contrast effects was found for all RRPSPS and SDQ ratings for female twins, but only for RRPSPS ratings at age 2 and SDQ ratings at age 4 in males. Although these findings suggest potential moderating effects of gender on rater contrast effects, the model did not take into account differences between DZ opposite-sex and DZ same-sex twin pairs. Therefore we did not exclude from further analysis male ratings which did not show evidence of contrast effects.

Step 2: Testing if contrast effects significantly differ between same-sex and opposite-sex twin pairs according to gender (twin correlation model with overall interaction effects)

In this series of models the moderating effects of gender composition on total contrast effects was tested. There was evidence to suggest that contrast effect parameters could be equated between same-sex males and opposite-sex males, and between same-sex females and opposite-sex female twins, for RRSPSC ratings at age 2 and SDQ ratings at age 7 (see Table 3). However, overall the evidence suggested that these interaction pathways were significantly different, and that these parameters would result in a significant deterioration of fit if equated.

Step 3: Testing moderators of contrast effects (twin correlation model with independent and moderator-dependent interaction effects)

Due to the lack of consistent evidence for equating same-sex and opposite-sex contrast effects parameters within genders, subsequent models testing for the contributory role of demographic factors on contrast effects specified four separate contrast effects (c_m, c_f, $c_{M\text{-}F}$, and $c_{F\text{-}M}$ Accordingly, these models tested the moderating effects of gender on rater contrast effects in general (independent component), and the interactive effects between gender and other moderators (family size, SES, and ethnicity) (moderator-dependent component).

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Moderating effects of gender—The independent components of contrast effect parameters were significant for same-sex pairs, across both genders (see Table 4). Although the effect was always larger for female same-sex pairs, they did not significantly differ from estimates for male same-sex pairs. Within DZ opposite-sex pairs a gender effect was observed: the independent component of the contrast effect parameter was not significant from male-to-female, but significant from female-to-male. The independent component of the contrast effect from females-to-males was larger than effects observed for same-sex pairs, but as confidence intervals (CIs) overlapped the magnitude of the difference was not statistically significant.

Interactive effect of gender and other socio-demographic moderators on contrast effects—When partitioning the contrast effect components moderated by SES or ethnicity, there was no evidence to suggest that these demographic variables contributed significantly to contrast effects, or that their effect was moderated by gender. Family size did not contribute to significant contrast effects in male same-sex pairs or DZ opposite-sex pairs. However, the family size dependent interaction parameter was small, but significant, in female same-sex pairs from age 4. Family size was a continuous variable, and the effect suggests that as family size increases the contrast effect parameter increases alongside. The positive value for these significant contrast effect parameters suggest that when combined with the negative contrast effect parameters found for independent contrast effect component, total contrast effects are reduced. Therefore parental ratings of twins from larger families are associated with smaller (overall) rater contrast effects.

DISCUSSION

This study employed a novel twin model to explicitly test for factors moderating contrast effects in parental ADHD ratings. This was achieved using a detailed examination of several socio-demographic variables in a large population-based twin sample, and partitioning the total contrast effects observed into independent and moderator-dependent components. Moreover, the inclusion of opposite-sex twins, allowed us to not only test the moderating and interactive role of gender across same-sex pairs, but also to test whether gender plays a differing role within opposite-sex twins.

The first main finding was that within same-sex twin pairs, parents contrast twins to a similar extent, regardless of whether they are female or male. Although gender did not moderate contrast effects (independent component) in same-sex pairs, there was an effect within opposite-sex pairs. Within opposite-sex pairs rater contrast parameters were significant only when the pathway was specified from females-to-males (from age 3), suggesting that males are being evaluated in relation to their female co-twin, who is considered the standard. The non-significant contrast effect parameters from males-to-females, suggests that when female members of an opposite-sex pair are being rated, their evaluation is being made independent of their male co-twin’s behavior. This was an unexpected finding, as the only previous study to distinguish contrast effects from male-to-female and female-to-male in opposite-sex pairs, observed a larger effect from males-to-females (Rietveld et al. 2003). The original direction of effect was hypothesized to stem from the assumption that these behaviors are more commonly associated with males, and so

Europe PMC Funders Author Manuscripts

males are considered the comparative benchmark, and females rated according to their male co-twin. However, if there is a greater awareness of normative standards for ADHD-related behaviors in males, one may also expect contrast effects to feature less in parental ratings of male versus female same-sex pairs. Yet findings from this study suggest that there are no differences in the magnitude of contrast effects across genders within same-sex pairs, in line with previous research (Rietveld et al. 2003; Simonoff et al. 1998). On the contrary, we speculate that our findings could be interpreted as suggesting that parents of opposite-sex pairs are more acutely aware of behavioural differences (actual and/or stereotypical). Accordingly, when female members of opposite-sex pairs are being rated by parents, their evaluation is being made independent of their male co-twins behavior (as suggested by our non-significant contrast effect parameter from males-to-females). This finding was consistent across ages and rating scales (with the exception of the first assessment point). However, it is important to note that this is the first attempt at a replication of the original finding and so further studies are needed to clarify the direction of the effect, although it may be that this observed discrepancy is based on sample differences or use of different assessment instruments.

Our second key finding was that family size did moderate contrast effects, in line with previous predictions (Carey 1986), and in the expected direction: parental ratings where family size was small was more likely to result in contrast effects. However, this effect was only observed for female same-sex pairs. The finding that contrast effects featured less in parental ratings of larger sized families might be explained by increased awareness of a broader range of child behaviors in these families, and a larger baseline for comparing rated behavior. Parents of smaller sized families, are more likely to have a smaller benchmark to make comparisons with, and are therefore more likely to directly compare twins. This finding is consistent with the hypothesis that teacher ratings of ADHD-related behaviors are free from contrast effects, as they have greater exposure to children and more objective standards of appropriate behavioral norms (Saudino et al. 2005; Hartman et al. 2007). The fact that the moderating effect of family size on contrast effects was only evident in females, could relate to ADHD-related behaviors being less commonly associated with females and consequently there being less clearly defined norms for these behaviors in girls; such that parents draw on a broader comparative group where possible (i.e. where they have a large family). However, the magnitude for the majority of the family-size dependent contrast effect parameters is similar across male and female same-sex pairs, and likely reaches significance in females due to a greater number of female same-sex twin pairs in our sample.. If we eliminate such cases the most robust gender difference for the effect of family-size on the interaction parameter is found at age 4, independent of rating scale.

Third, we replicate previous findings that not all parent rating scales are equally susceptible to rater bias, and that the Conners' Parent Rating Scale is particularly robust against contrast effects, confirming previous observations that particularly implicate short general rating scales (Price et al. 2005; Price et al. 2001), rather than longer more detailed scales such as DSM-IV checklists (Kuntsi et al. 2005). This may be attributable to the broader scoring range and/or more detailed and precise nature of the items being rated, compared to the other scales examined. This is consistent with the observation that the SWAN (Strengths and Weaknesses of ADHD Symptoms and Normal Behaviors) (Swanson et al. 2006), which uses similar detailed questions and a 7-point scoring range to measure both positive and negative behaviors, seems immune to contrast effects (Hay et al. 2007; Polderman et al. 2007). Another possibility is that the Conners' Parent Rating Scale with 18 items that parallel current DSM criteria may be measuring a partially different behavioral trait compared to the other shorter rating scales used in this study. However, it may also be the case that the lack of contrast effects in ratings from the Conners' Parent Rating Scale obtained at ages 8 and 12, reflect an age-effect: that contrast effects are more prevalent in ratings of younger, rather than older children. We were unable to test this directly, as we did not have ratings from this scale at younger ages.

The other socio-demographic factors that we investigated (ethnicity and SES) did not contribute to parental rater contrast effects. Despite methodological and sample differences, our finding that SES did not contribute to contrast effects is in line with previous findings in a non-twin sample (Saudino et al. 2004). In order to potentially control for their effects, or improve rating scales, further research is needed to clarify the origins of rater contrast effects. For example, societal factors may be important and variations in contrast effects in parental ratings from collectivist versus individualistic societies could be examined (Saudino et al. 2004).

A limitation of this study is that we did not have information on parental ethnicity. Consequently we used childrens' ethnicity as a marker, albeit less accurate, of parental ethnicity, to test for cultural differences in parents' tendency to contrast twins. A further limitation is that our family size variable was collated from information when twins were aged 18 months, and could not have taken into account additional children previously or subsequently residing in the household. Taking these factors into account, replication is needed before firm conclusions can be made.

There is only one other study that we are aware of to test whether parental demographics contributed to contrast effects in ADHD-related ratings, and this was carried out using correlational analysis between difference scores on ratings and parental demographics in a small non-twin sibling sample (Saudino et al. 2004). Ideally contrast effects can be formally tested using structural equation modeling, and our study employed a particularly novel methodology to tease apart contrast effects to determine underlying sources of parental contrast effects. An important consideration for future research is sample size, as there is limited power to detect contrast effects in small twin samples, especially when genetic dominance effects may be present (Rietveld et al. 2003), as has been reported in a meta-analysis of ADHD behaviors (Burt 2009). Consequently, our large sample size was a major strength in this study.

Contrast effects in parental ratings of ADHD symptoms clearly need to be acknowledged as a potential bias in quantitative genetic research, which may have contributed to over-estimated heritability estimates, which is an important consideration for phenotype selection in molecular investigations. In addition, it may be the case that inflated heritability has contributed to less attention being directed at environmental factors underlying ADHD presentation. Our study identified, from the socio-demographic variables investigated, family size as a significant contributor to contrast effects. This finding is consistent with the

Europe PMC Funders Author Manuscripts

view that contrast effects reflect parental biases, rather than actual behavioral differences, as phenotypic differences in the presentation of ADHD-related behaviors are unlikely to vary according to family size. Overall, this and previous studies (Thapar et al. 2000; Sherman et al. 1997) support the use of multiple informants in studies on ADHD. Our findings further indicate that selection of rating scale does matter: studies should also endeavor to use measurement scales that are less susceptible to contrast effects. In addition, research directed at obtaining more objective cognitive, metabolic or neurological ADHD markers will help overcome the reliance on ADHD symptom scales and the biases they may be associated with.

Acknowledgments

We are indebted to participants of the Twins Early Development Study (TEDS) for making the study possible. The authors have no financial relationships to disclose. TEDS is funded by MRC grant G0500079.

REFERENCES

Burt SA. Rethinking environmental contributions to child and adolescent psychopathology: a meta-analysis of shared environmental influences. Psychol Bull. 2009; 135(4):608–637. [PubMed: 19586164]

Carey G. Sibling imitation and contrast effects. Behav Genet. 1986; 16(3):319–341. [PubMed: 3753368]

Conners CK, Sitarenios G, Parker JD, Epstein JN. The revised Conners' Parent Rating Scale (CPRS-R): factor structure, reliability, and criterion validity. J Abnorm Child Psychol. 1998; 26(4):257–268. [PubMed: 9700518]

Faraone SV, Perlis RH, Doyle AE, Smoller JW, Goralnick JJ, Holmgren MA, Sklar P. Molecular genetics of attention-deficit/hyperactivity disorder. Biological Psychiatry. 2005; 57(11):1313–1323. [PubMed: 15950004]

Freitag CM, Rohde LA, Lempp T, Romanos M. Phenotypic and measurement influences on heritability estimates in childhood ADHD. European Child & Adolescent Psychiatry. 2010; 19(3): 311–323. [PubMed: 20213230]

Goodman R. The Strengths and Difficulties Questionnaire: a research note. J Child Psychol Psychiatry. 1997; 38(5):581–586. [PubMed: 9255702]

Greven CU, Rijsdijk FV, Plomin R. A twin study of ADHD symptoms in early adolescence: hyperactivity-impulsivity and inattentiveness show substantial genetic overlap but also genetic specificity. J Abnorm Child Psychol. 2011; 39(2):265–275. [PubMed: 21336711]

Hartman CA, Rhee SH, Willcutt EG, Pennington BF. Modeling rater disagreement for ADHD: are parents or teachers biased? J Abnorm Child Psychol. 2007; 35(4):536–542. [PubMed: 17333362]

Hay DA, Bennett KS, Levy F, Sergeant J, Swanson J. A twin study of attention-deficit/hyperactivity disorder dimensions rated by the strengths and weaknesses of ADHD-symptoms and normal-behavior (SWAN) scale. Biol Psychiatry. 2007; 61(5):700–705. [PubMed: 16962074]

Hogg, C.; Rutter, M.; Richman, N. Emotional and behaviural problems in children. NFER-Nelson; Windsor: 1997.

Hudziak JJ, Derks EM, Althoff RR, Copeland W, Boomsma DI. The genetic and environmental contributions to oppositional defiant behavior: a multi-informant twin study. J Am Acad Child Adolesc Psychiatry. 2005; 44(9):907–914. [PubMed: 16113619]

Kumarapeli P, Stepaniuk R, de Lusignan S, Williams R, Rowlands G. Ethnicity recording in general practice computer systems. J Public Health (Oxf). 2006; 28(3):283–287. [PubMed: 16840765]

Kuntsi J, Eley TC, Taylor A, Hughes C, Asherson P, Caspi A, Moffitt TE. Co-occurrence of ADHD and low IQ has genetic origins. Am J Med Genet B Neuropsychiatr Genet. 2004; 124B(1):41–47. [PubMed: 14681911]

Kuntsi J, Rijsdijk F, Ronald A, Asherson P, Plomin R. Genetic influences on the stability of attention-deficit/hyperactivity disorder symptoms from early to middle childhood. Biol Psychiatry. 2005; 57(6):647–654. [PubMed: 15780852]

Martin N, Scourfield J, McGuffin P. Observer effects and heritability of childhood attention-deficit hyperactivity disorder symptoms. Br J Psychiatry. 2002; 180:260–265. [PubMed: 11872519]

McLoughlin G, Ronald A, Kuntsi J, Asherson P, Plomin R. Genetic support for the dual nature of attention deficit hyperactivity disorder: Substantial genetic overlap between the inattentive and hyperactive-impulsive components. Journal of Abnormal Child Psychology. 2007; 35(6):999–1008. [PubMed: 17690977]

Merwood A, Greven CU, Larsson H, Price TS, Asherson P. Different heritabilities but shared aetiological influences for parent, teacher and self ratings of ADHD symptoms: an adolescent twin study. Psychological Medicine. Submitted.

Neale, MC. Mx. Department of Psychiatry, Medical College of Virginia; Richmond: 1997.

Pike A, Iervolino AC, Eley TC, Price TS, Plomin R. Environmental risk and young children's cognitive and behavioral development. International Journal of Behavioral Development. 2006; 30(1):55–66.

Plomin, R.; DeFries, JC.; McClearn, GE.; McGuffin, P. Behavioral genetics. Worth Publishers; New York: 2008.

Polderman TJ, Derks EM, Hudziak JJ, Verhulst FC, Posthuma D, Boomsma DI. Across the continuum of attention skills: a twin study of the SWAN ADHD rating scale. J Child Psychol Psychiatry. 2007; 48(11):1080–1087. [PubMed: 17995483]

Price TS, Freeman B, Craig I, Petrill SA, Ebersole L, Plomin R. Infant zygosity can be assigned by parental report questionnaire data. Twin Res. 2000; 3(3):129–133. [PubMed: 11035484]

Price TS, Simonoff E, Asherson P, Curran S, Kuntsi J, Waldman I, Plomin R. Continuity and change in preschool ADHD symptoms: longitudinal genetic analysis with contrast effects. Behav Genet. 2005; 35(2):121–132. [PubMed: 15685426]

Price TS, Simonoff E, Waldman I, Asherson P, Plomin R. Hyperactivity in preschool children is highly heritable. J Am Acad Child Adolesc Psychiatry. 2001; 40(12):1362–1364. [PubMed: 11765277]

Rietveld MJ, Hudziak JJ, Bartels M, van Beijsterveldt CE, Boomsma DI. Heritability of attention problems in children: I. cross-sectional results from a study of twins, age 3-12 years. Am J Med Genet B Neuropsychiatr Genet. 2003; 117B(1):102–113. [PubMed: 12555244]

Rietveld MJ, Posthuma D, Dolan CV, Boomsma DI. ADHD: sibling interaction or dominance: an evaluation of statistical power. Behav Genet. 2003; 33(3):247–255. [PubMed: 12837015]

Saudino KJ, Cherny SS, Plomin R. Parent ratings of temperament in twins: explaining the 'too low' DZ correlations. Twin Res. 2000; 3(4):224–233. [PubMed: 11463143]

Saudino KJ, Ronald A, Plomin R. The etiology of behavior problems in 7-year-old twins: substantial genetic influence and negligible shared environmental influence for parent ratings and ratings by same and different teachers. J Abnorm Child Psychol. 2005; 33(1):113–130. [PubMed: 15759595]

Simonoff E, Pickles A, Hervas A, Silberg JL, Rutter M, Eaves L. Genetic influences on childhood hyperactivity: contrast effects imply parental rating bias, not sibling interaction. Psychol Med. 1998; 28(4):825–837. [PubMed: 9723139]

Saudino KJ, Wertz AE, Gagne JR, Chawla S. Night and day: are siblings as different in temperament as parents say they are? J Pers Soc Psychol. 2004; 87(5):698–706. [PubMed: 15535780]

Sherman DK, McGue MK, Iacono WG. Twin concordance for attention deficit hyperactivity disorder: a comparison of teachers' and mothers' reports. Am J Psychiatry. 1997; 154(4):532–535. [PubMed: 9090341]

StataCorporation. Stata statistical software. Stata Corporation; College Station: 1997.

Swanson, JM.; Schuck, S.; Mann, M.; Carlson, C.; Hartman, K.; Sergeant, JA.; Clevinger, W.; Wasdell, M.; McCleary, R. Categorical and dimensional definitions and evaluations of symptoms of ADHD: The SNAP and SWAN Rating Scales. 2006. http://wwwADHDnet InRetrieved December 2010 from

Thapar A, Harrington R, Ross K, McGuffin P. Does the Definition of ADHD Affect Heritability? J Am Acad Child Adolesc Psychiatry. 2000; 39(12):1528–1536. [PubMed: 11128330]

Europe PMC Funders Author Manuscripts

Trouton A, Spinath FM, Plomin R. Twins early development study (TEDS): a multivariate, longitudinal genetic investigation of language, cognition and behavior problems in childhood. Twin Res. 2002; 5(5):444–448. [PubMed: 12537874]

van Beijsterveldt CE, Verhulst FC, Molenaar PC, Boomsma DI. The genetic basis of problem behavior in 5-year-old dutch twin pairs. Behav Genet. 2004; 34(3):229–242. [PubMed: 14990864]

Vierikko E, Pulkkinen L, Kaprio J, Rose RJ. Genetic and environmental influences on the relationship between aggression and hyperactivity-impulsivity as rated by teachers and parents. Twin Res. 2004; 7(3):261–274. [PubMed: 15193171]

Wood AC, Buitelaar J, Rijsdijk F, Asherson P, Kuntsi J. Rethinking shared environment as a source of variance underlying attention-deficit/hyperactivity disorder symptoms: comment on Burt (2009). Psychol Bull. 2010; 136(3):331–340. [PubMed: 20438137]

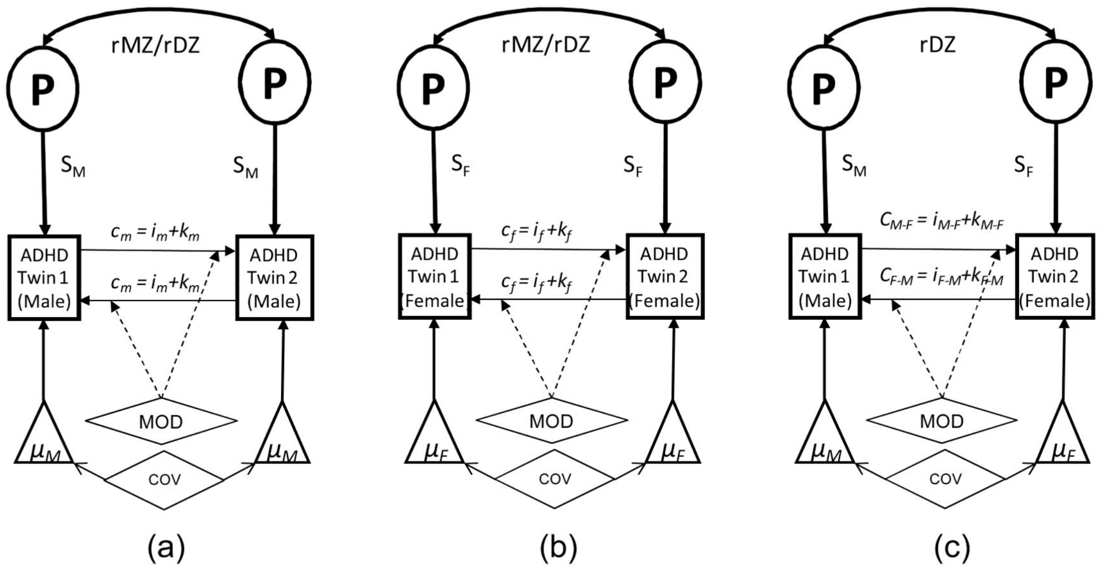

Figure 1. The variance-covariance model of MZ DZ twin pairs by varying gender composition

Note: The variance-covariance model of MZ, DZ same-sex pairs for males (a), females (b) and DZ opposite-sex pairs (c). P is the phenotypic variation. S is the standard deviation of the ADHD symptom scores, assumed to be similar across twin order and zygosity groups, but differ across gender (S_M for males, S_F for females). The correlation between the scores is estimated separately for MZ and DZ pairs, but specified to be the same across gender and across DZ same-sex and DZ opposite-sex pairs. The reciprocal causal paths between the phenotypes (*c*) are composed of a part independent of the specifically modelled moderators, indicated by *i*, and a moderator specific part (*k*). These effects differ according to gender composition of pairs, and are modeled separately for males ($_m$), females ($_f$), and for males-to-females ($_{M\text{-}F}$) and females to males ($_{F\text{-}M}$). MOD are definition variables modeling the moderator effect of ethnicity, SES and family size on the interaction terms and COV are definition variables modeling the effects of covariates (age) in the model of the means.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

Table 1

Number of twin pairs by zygosity and gender for each socio-demographic variable

	Additional Children											Ethnicity		SES
	None	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**White**	**Ethnic Minority**	**SES**
MZM	656	527	188	64	9	1	1	1	1	0	0	1342	102	1327
MZF	730	592	217	54	18	0	3	1	1	0	0	1482	129	1467
DZSSM	689	490	187	42	17	5	1	0	0	0	0	1340	88	1334
DZSSF	696	466	193	73	15	7	0	0	0	0	0	1348	100	1341
DZOS	1396	917	354	122	29	4	2	0	0	0	1	2623	194	2590
Total	4167	2992	1139	355	88	17	7	2	2	0	1	8135	613	8059

Note: MZM: MZ males; MZF: monozygotic females; DZSSM: dizygotic same-sex males; DZSSF: dizygotic same-sex females; DZOS- dizygotic opposite sex.

Table 2

Twin correlations and variance estimates by zygosity: testing zygosity differences in variances

	Correlations		Variances				Test if vMZ = vDZ (χ^2 (2 df))[d]	
ADHD Scale (age)	**rMz**	**rDZ[a]**	**vMZM**	**vDZM[b]**	**vMZF**	**vDZF[c]**	**Males**	**Females**
RRPSPC (2)	0.65	0.18	3.21	3.85	3.22	3.93	14.00 ($p < 0.001$)	18.53 ($p < 0.001$)
RRPSPC (3)	0.64	0.03	3.64	3.91	3.30	3.91	2.22 ($p = 0.14$)	13.78 ($p < 0.001$)
RRPSPC (4)	0.58	−0.02	3.55	3.84	3.03	3.69	3.93 ($p = 0.05$)	25.89 ($p < 0.001$)
SDQ (4)	0.50	−0.07	5.04	5.67	4.32	5.51	8.99 ($p = 0.003$)	41.09 ($p < 0.001$)
SDQ (7)	0.57	−0.03	6.49	6.69	5.36	6.42	0.63 ($p = 0.43$)	24.64 ($p < 0.001$)
SDQ (12)	0.74	0.18	5.10	5.48	3.80	4.92	2.63 ($p = 0.11$)	37.94 ($p < 0.001$)
CPRS-R (8)	0.86	0.47	0.34	0.36	0.32	0.35	1.70 ($p = 0.19$)	3.98 ($p = 0.05$)
CPRS-R (12)	0.86	0.45	0.62	0.41	0.38	0.40	1.94 ($p = 0.16$)	1.69 ($p = 0.16$)

Note: rMZ and rDZ refer to twin correlations; vMZM and vMZF refer to variance estimates; vDZM and vDZF refer to variance estimates.

RRPSPC refers to Revised Parent Rutter Scale for Pre-School Children; SDQ refers to Strengths and Difficulties Questionnaire; CPRS-R refers to the Revised Conners' Parent Rating Scale; CPRS-R ratings were positively skewed and therefore normalized for analysis.

[a] includes DZOS

[b] includes DZOSM

[c] includes DZOSF

[d] Compared a model where variances are constrained across zygosity (2 df)

Europe PMC Funders Author Manuscripts

Table 3

Contrast effect parameters between same-sex and opposite-sex pairs, by gender

	Males			Females		
ADHD Scale (age)	c_m	$c_{M\text{-}F}$	**Test if** $c_m = c_{M\text{-}F}$ (χ^2 (1 df))[a]	c_f	$c_{M\text{-}F}$	**Test if** $c_f = c_{F\text{-}M}$ (χ^2 (1 df))[a]
RRPSPC (2)	−0.19	−0.10	2.31 (p = 0.13)	−0.20	−0.23	0.38 (p = 0.54)
RRPSPC (3)	−0.14	0.08	7.17 (p = 0.007)	−0.16	−0.47	9.74 (p =0.002)
RRPSPC (4)	−0.18	0.008	6.02 (p = 0.01)	−0.19	−0.41	6.88 (p = 0.009)
SDQ (4)	−0.22	0.02	7.31 (p = 0.007)	−0.22	−0.48	7.06 (p = 0.008)
SDQ (7)	−0.11	0.03	1.60 (p = 0.21)	−0.13	−0.32	1.75 (p = 0.19)
SDQ (12)	−0.27	−0.15	12.66 (p < 0.000)	−0.25	−0.52	37.58 (p < 0.000)

Note: c_m refer to contrast effect parameter between males; c_f refer to contrast effect parameter between females; $c_{M\text{-}F}$ refer to contrast effect parameter from male-to-female; $c_{M\text{-}F}$ refer to contrast effect parameter from female-to-male

[a]Compared a model where variances are constrained across SS and OS pairs, for males and females separately (1 df); RRPSPC refers to Revised Parent Rutter Scale for Pre-School Children; SDQ refers to Strengths and Difficulties Questionnaire; CPRS-R ratings were not included in analysis as they showed no significant variance differences by zygosity in initial stage of analysis, suggesting the absence of contrast effects;

Europe PMC Funders Author Manuscripts

Table 4

Contrast effect parameters (and 99% confidence intervals CIs) by gender composition of twin pairs- decomposed into independent and moderator-dependent components

	ADHD rating Scale (age)					
	RRPSPC (2)	**RRPSPC (3)**	**RRPSPC (4)**	**SDQ (4)**	**SDQ (7)**	**SDQ (12)**
Independent						
cm	**−0.19 (−0.42 to −0.05)**	**−0.14 (−0.25 to −0.04)**	**−0.19 (−0.32 to −0.10)**	**−0.25 (−0.36 to −0.15)**	**−0.17 (−0.26 to −0.18)**	**−0.27 (−0.44 to −0.15)**
cf	**−0.19 (−0.42 to −0.06)**	**−0.18 (−0.29 to −0.08)**	**−0.24 (−0.37 to −0.15)**	**−0.31 (−0.42 to −0.21)**	**−0.18 (−0.27 to −0.09)**	**−0.29 (−0.46 to −0.16)**
cM-F	−0.16 (−0.43 to 0.13)	0.04 (−0.24 to 0.23)	−0.12 (−0.32 to 0.07)	−0.12 (−0.35 to 0.06)	0.02 (−0.19 to 0.17)	−0.16 (−0.36 to 0.02)
cF-M	−0.23 (−0.46 to 0.04)	**−0.48 (−0.68 to −0.11)**	**−0.36 (−0.57 to −0.09)**	**−0.42 (−0.63 to −0.12)**	**−0.37 (−0.59 to −0.08)**	**−0.50 (−0.68 to −0.32)**
Dependent						
Family size						
cm	0.005 (−0.04 to 0.05)	−0.0008 (−0.04 to 0.04)	−0.0001 (−0.04 to 0.04)	0.01 (−0.02 to 0.05)	0.04 (−0.01 to 0.07)	0.03 (−0.01 to 0.06)
cf	0.002 (−0.03 to 0.04)	0.01 (−0.02 to 0.05)	**0.06 (0.03 to 0.09)**	**0.08 (0.05 to 0.11)**	**0.03 (0.01 to 0.06)**	**0.04 (0.01 to 0.07)**
cM-F	0.07 (−0.06 to 0.19)	0.02 (−0.18 to 0.18)	0.13 (−0.002 to 0.25)	0.11 (−0.02 to 0.24)	0.02 (−0.13 to 0.15)	0.02 (−0.06 to 0.09)
cF-M	0.001 (−0.11 to 0.12)	0.02 (−0.14 to 0.21)	−0.05 (−0.17 to 0.08)	−0.04 (−0.17 to 0.10)	0.02 (−0.02 to 0.18)	−0.03 (−0.11 to 0.06)
SES						
cm	−0.02 (−0.07 to 0.02)	−0.02 (−0.07 to 0.02)	−0.02 (−0.06 to 0.03)	−0.01 (−0.06 to 0.03)	0.02 (−0.02 to 0.06)	−0.001 (−0.04 to 0.04)
cf	0.004 (−0.04 to 0.05)	0.03 (−0.01 to 0.08)	0.006 (−0.03 to 0.05)	0.02 (−0.02 to 0.06)	0.02 (−0.01 to 0.06)	0.002 (−0.04 to 0.04)
cM-F	−0.03 (−0.19 to 0.15)	0.007 (−0.16 to 0.18)	−0.07 (−0.25 to 0.11)	0.05 (−0.11 to 0.19)	0.15 (−0.01 to 0.28)	−0.04 (−0.13 to 0.06)
cF-M	−0.004 (−0.17 to 0.16)	−0.01 (−0.19 to 0.17)	0.06 (−0.15 to 0.26)	−0.07 (−0.24 to 0.11)	−0.06 (−0.21 to 0.12)	0.02 (−0.11 to 0.14)
Ethnicity						
cm	−0.04 (−0.20 to 0.12)	−0.06 (−0.21 to 0.10)	0.05 (−0.10 to 0.19)	0.05 (−0.08 to 0.19)	0.08 (−0.05 to 0.21)	0.07 (−0.07 to 0.21)
cf	−0.07 (−0.21 to 0.07)	0.09 (−0.06 to 0.24)	−0.02 (−0.14 to 0.11)	−0.006 (−0.13 to 0.12)	0.08 (−0.19 to 0.03)	−0.03 (−0.15 to 0.09)
cM-F	0.04 (−0.41 to 0.46)	−0.09 (−0.74 to 0.47)	−0.39 (−0.82 to 0.44)	0.22 (−0.57 to 0.57)	−0.01 (−0.71 to 0.34)	−0.08 (−0.36 to 0.16)
cF-M	0.10 (−0.31 to 0.55)	0.32 (−0.28 to 0.99)	0.50 (−0.47 to 0.95)	−0.13 (−0.54 to 0.69)	−0.05 (−0.43 to 0.63)	0.08 (−0.24 to 0.40)

Note: c_m refer to contrast effect parameter between males; c_f refer to contrast effect parameter between females; $c_{M\text{-}F}$ refer to contrast effect parameter from male-to-female; $c_{M\text{-}F}$ refer to contrast effect parameter from female-to-male; significant contrast effects (i.e. 99% CI not overlapping with the value zero) are in bold; RRPSPC refers to Revised Parent Rutter Scale for Pre-School Children; SDQ refers to Strengths and Difficulties Questionnaire; CPRS-R ratings were not included in analysis as they showed no significant variance differences by zygosity in initial stage of analysis, suggesting the absence of contrast effects;